

# Lending club case analysis

By Bideepta Basu and Rajeev Raga

---

# Problem Statement

---

- This Lending Loan Case Study is related to risk and analysis of providing loan to customers.
- Depending on the applicant's profile, company/bank need to decide the loan approval and rejection.
- Main risks that are associated in making loan decision are based on defaulters list of the applicants.
- Identifying multiple patterns from the applicant's data will help the company to take appropriate action against loan applications.
- Exploratory Data Analysis (EDA) plays major role in deriving the required insights for multiple variables of dataset.
- Identification of defaulters from the given applicants data will allow company to do the risk assessment



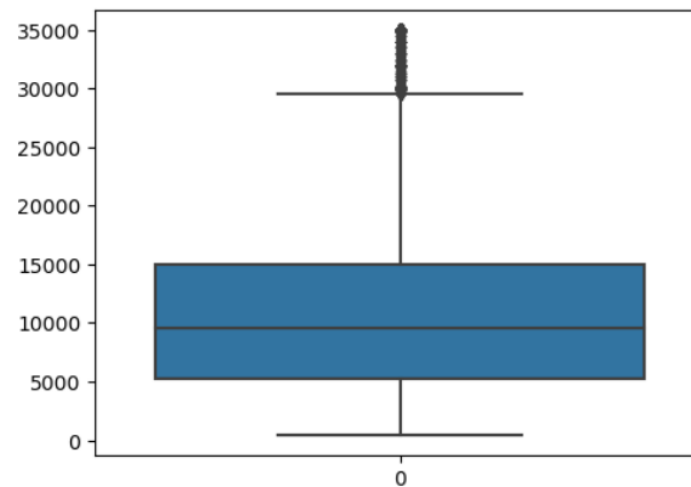
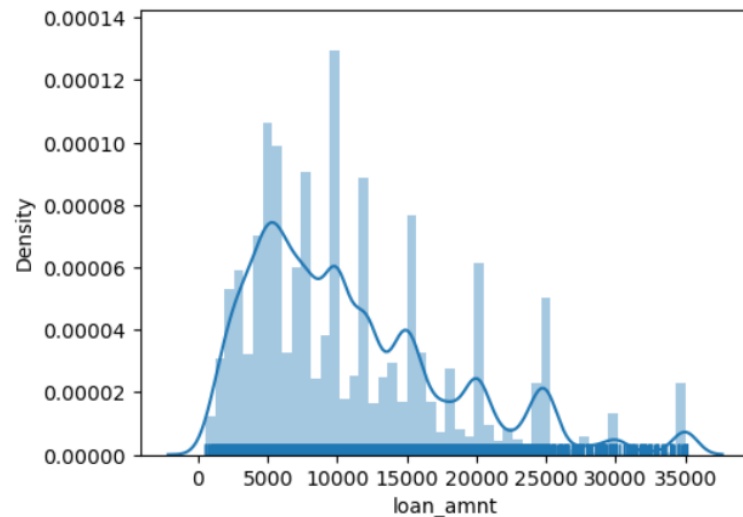
# Problem Approach

---

- Importing Required Libraries
- Importing the data from Loan.csv
- Validating the shape of the data imported
- Identifying the missing/null values in the dataset and dealing with them so that there should not be any issues to derive insights
- Checking if there are any outliers exist for the variables in dataset
- After data validation is completed then performed Univariate analysis on a particular column
- Performed Bivariate analysis with 2 different columns at a time to get more insights

# Univariant Analysis

## 1) Analysis on Loan Amount:

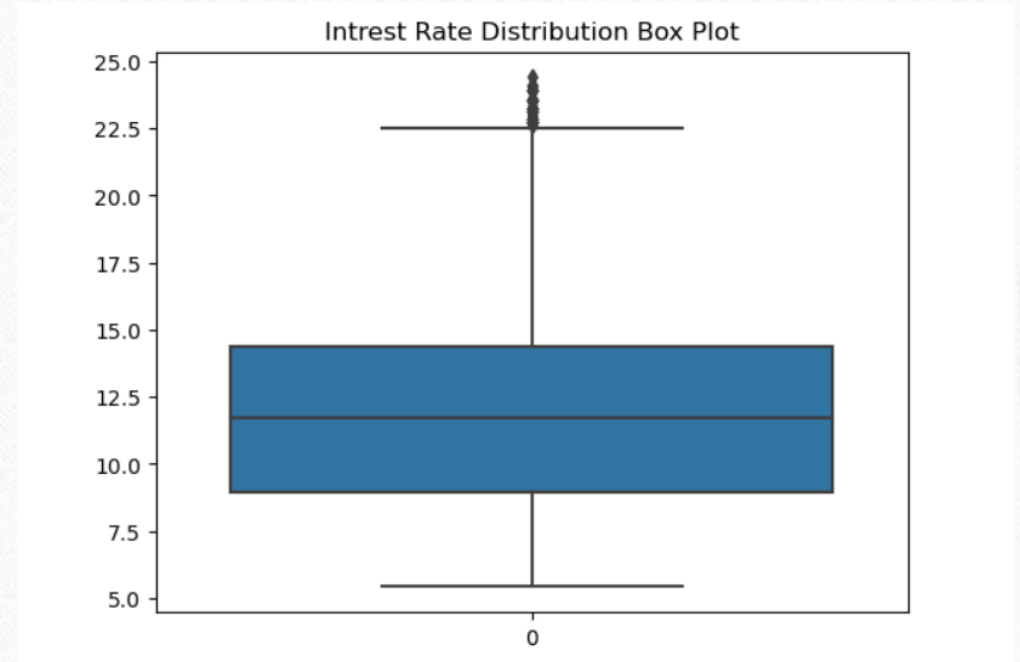


Above two graphs states that average loan amount taken by a customer is 10000 with same median value

# Cont..

## 2. Analysis on Interest Rate:

1. From this box plot it is clear that most of the loans taken are falls under the interest rate window 9 to 14.75 approx. which signifies that there is a high chance of customer getting default with higher interest rates.
2. Here the highest interest rate is 24.4 %

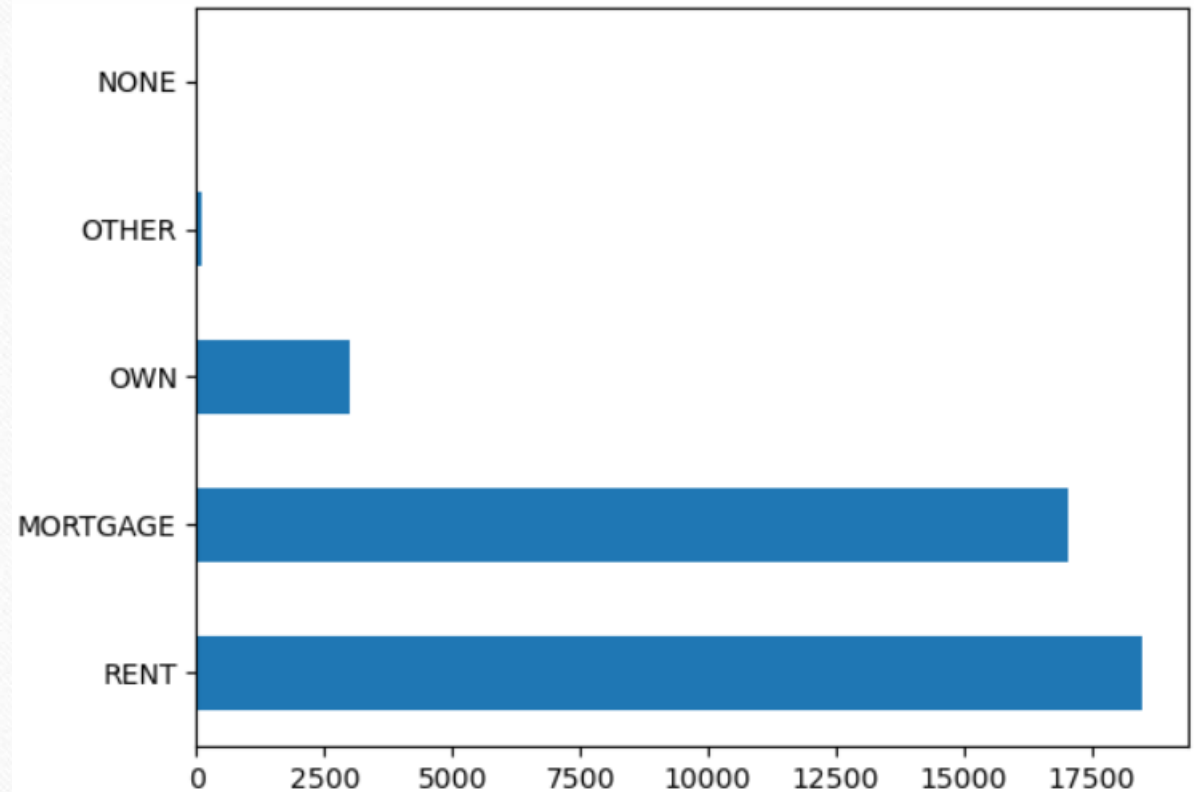




# Cont..

## 3. Analysis on Home Ownership:

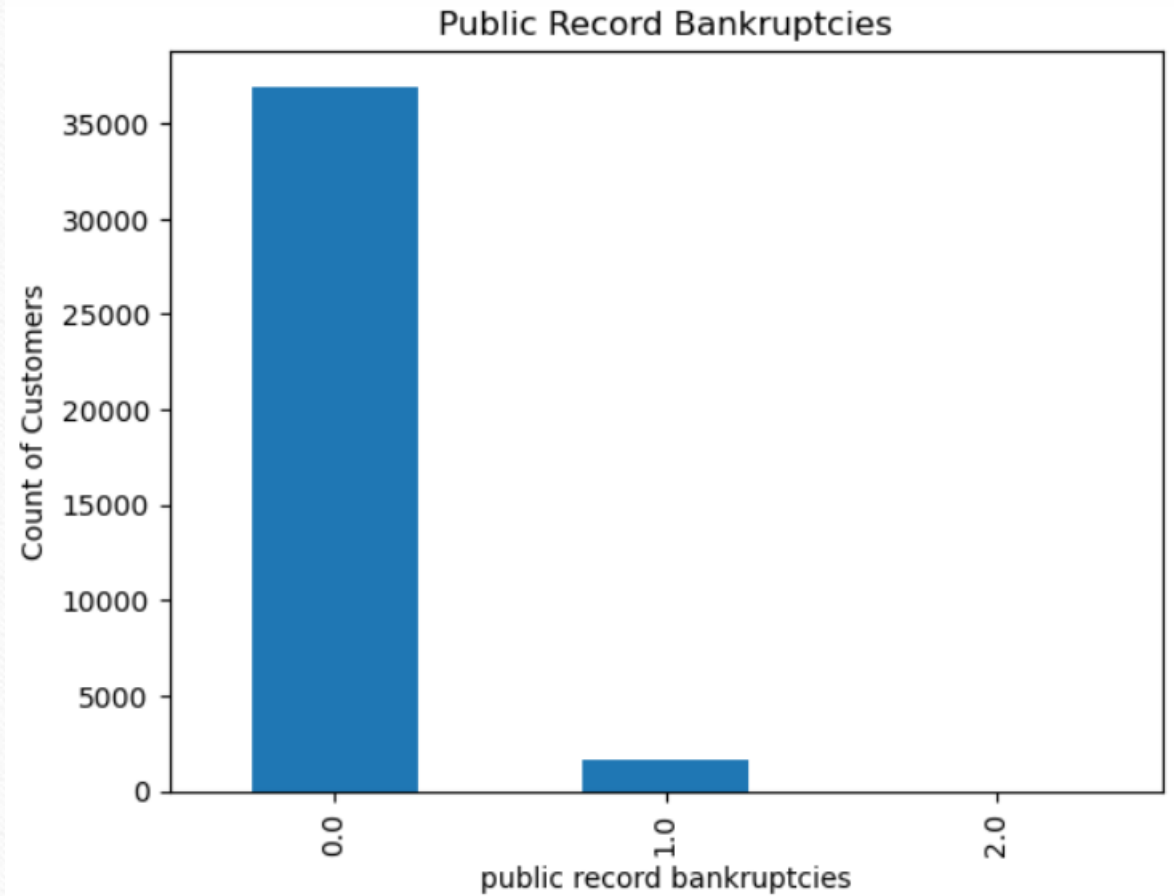
1. Here the more number of loans are taken by customers who are staying in 'Rented' place which can be a significant factor of customer getting default as they can vacate the Home without prior information.
2. There is a minor difference between 'MORTGAGE and RENT' in the customers who took the loan
3. The count of customers who are staying in their 'OWN' property is very low



# Cont..

## 4. Analysis on Public Record Bankruptcies(PRB):

1. Here the customers with 0 PRB is high and the risk involved in defaulting the customers under this category is less
2. There are very less customers whose PRB is 1

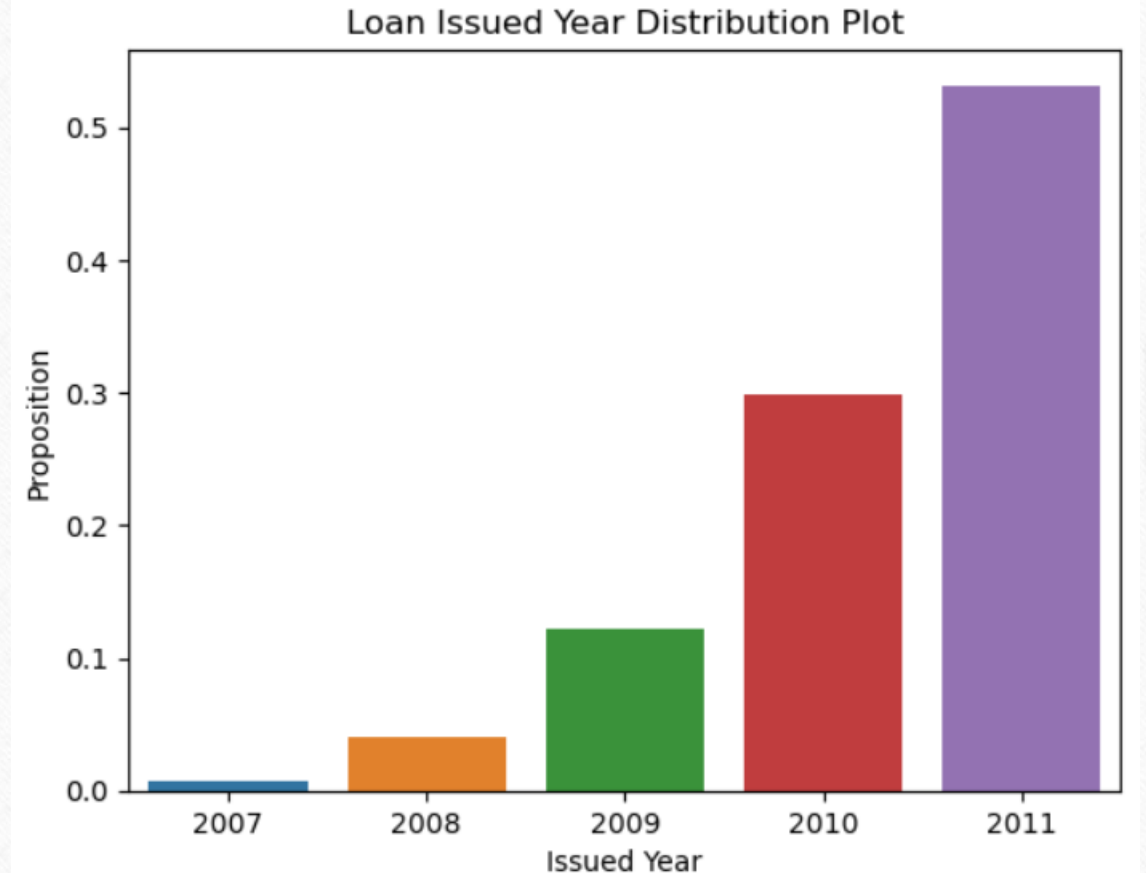


# Cont..

## 4. Analysis on Issue Date:

### 4.1 Year:

1. The maximum loans dispersed in the year 2011
2. Here we can clearly notice that there is a huge growth in loan dispersal year-on-year basis.

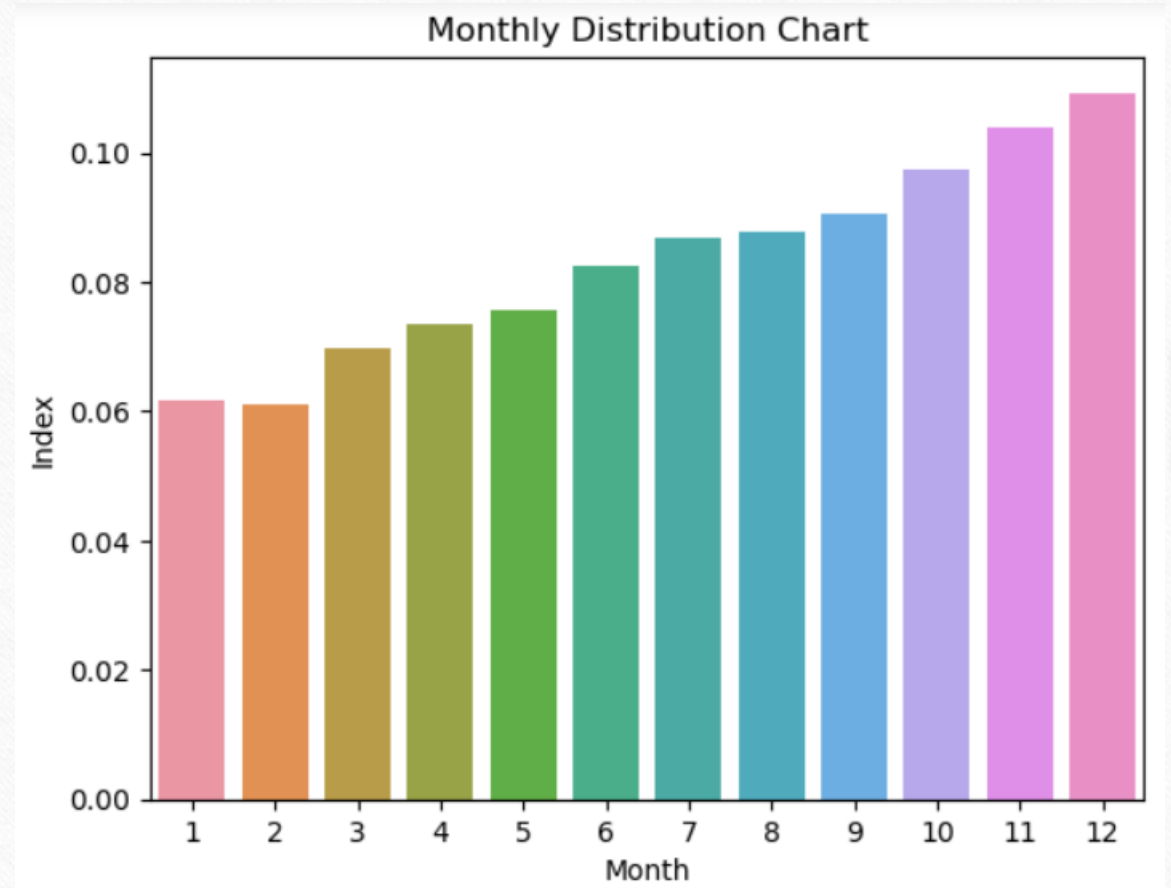




# Cont..

## 4.2 Month:

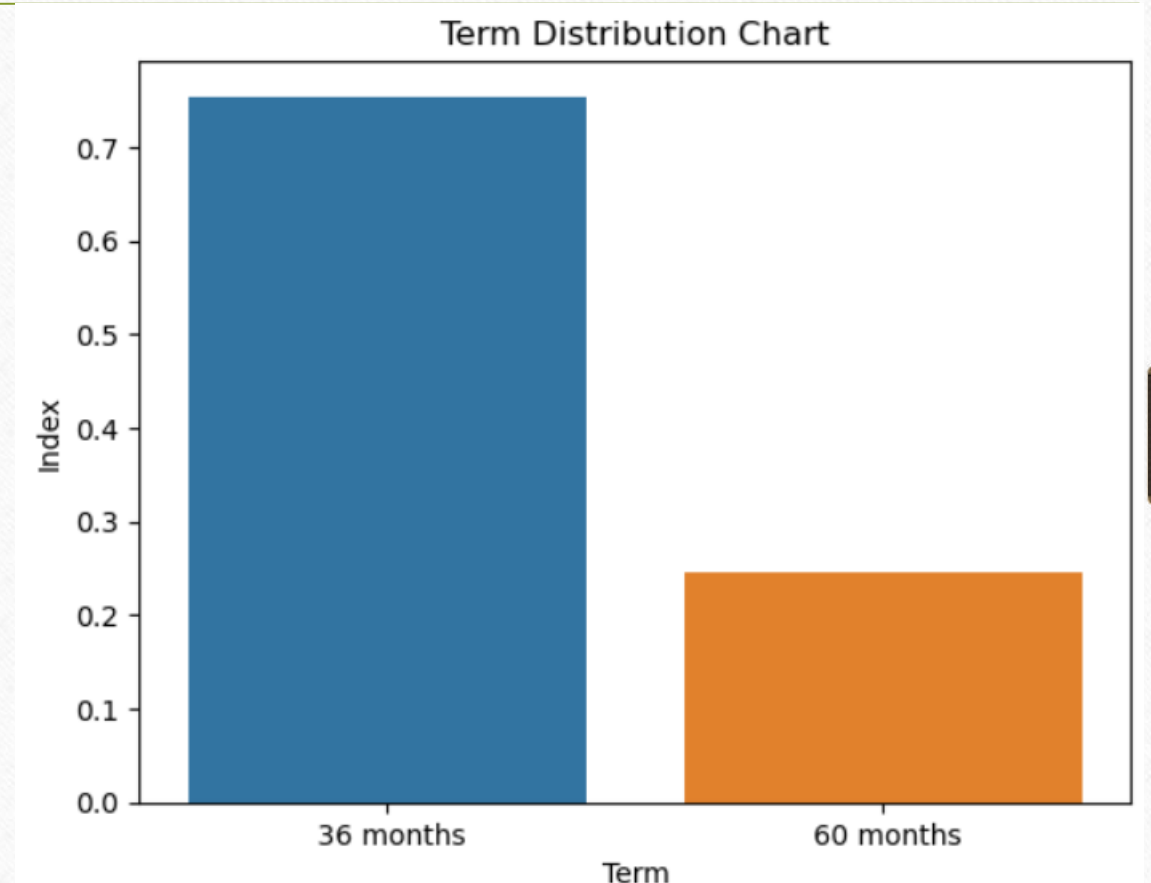
1. The maximum loans dispersed in the month of December
2. Here we can clearly notice that the average amount of loans are dispersed in between July and September months.



# Cont..

## 5. Analysis on Term:

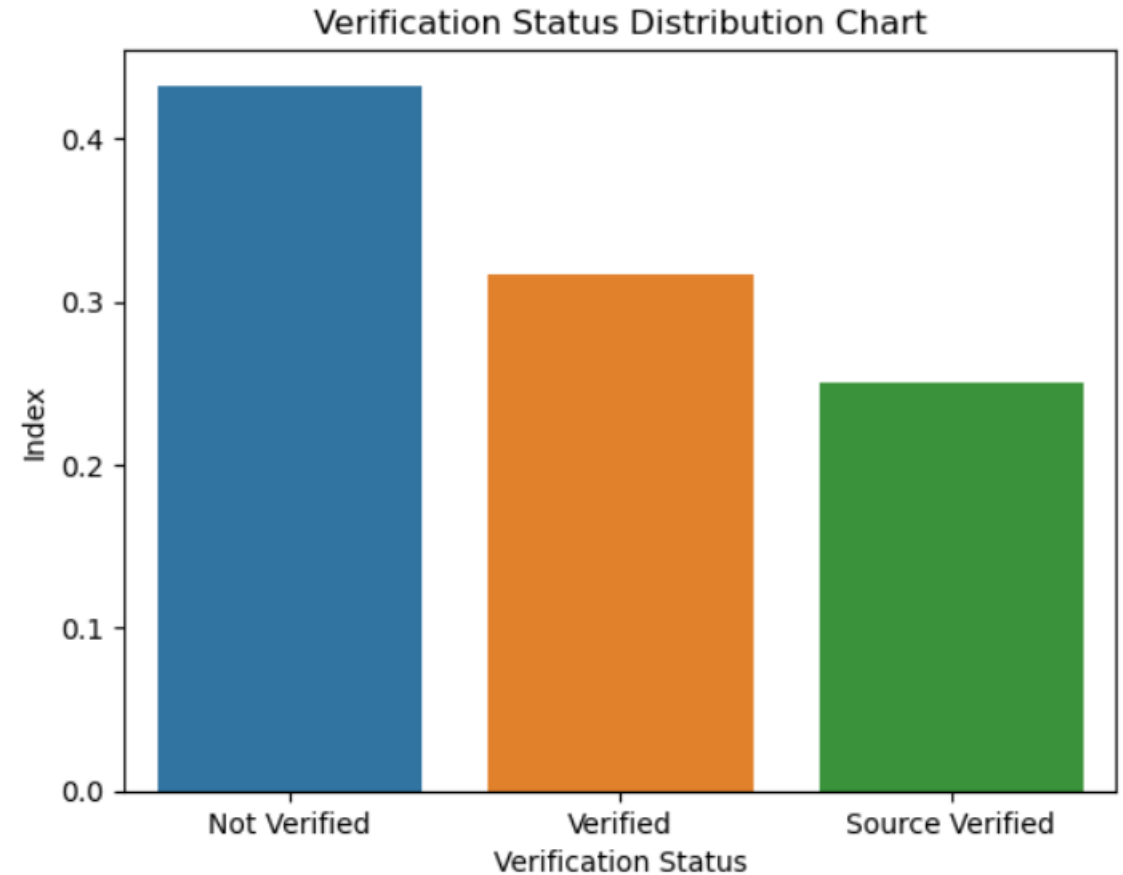
1. Here we can notice that maximum number of loans dispersed are for 36 Months.
2. The loans dispersed for the duration 60 Months is less.
3. Around 76% of customers took the loan for 36 months



# Cont..

## 6. Analysis on Verification Status:

1. Here we can notice that maximum number of loans dispersed are having verification status as 'NOT VERIFIED' which signifies the major risk involved in defaulting of loans.

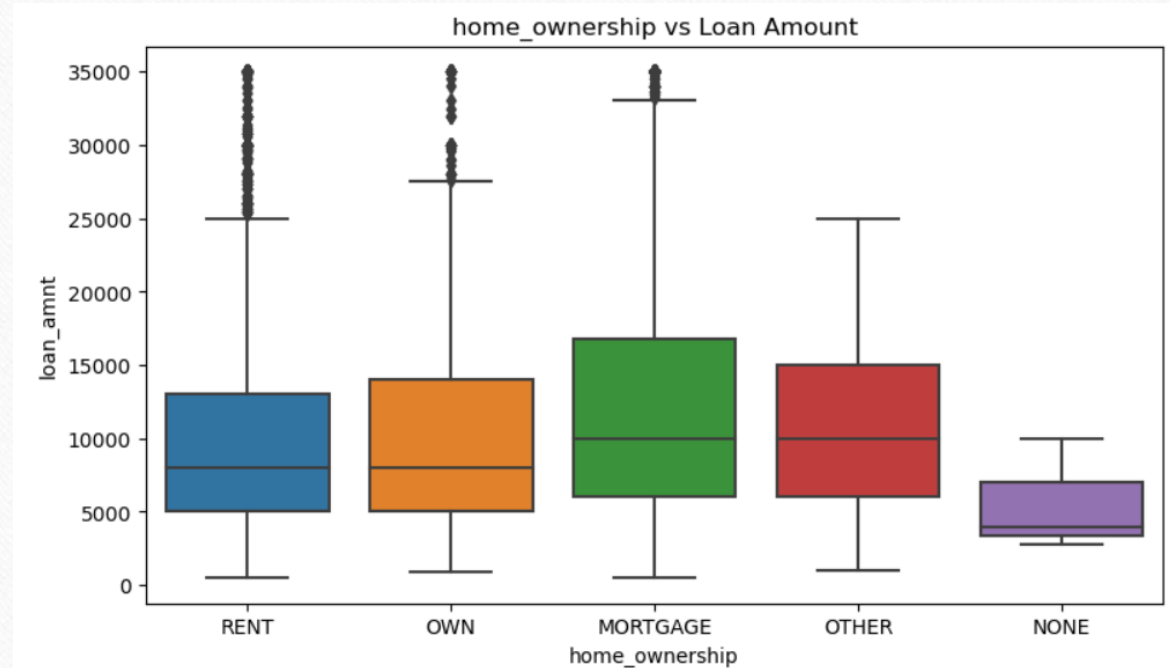




# Segmented Univariant Analysis

## 1) Analysis on Loan Amount with Home Ownership:

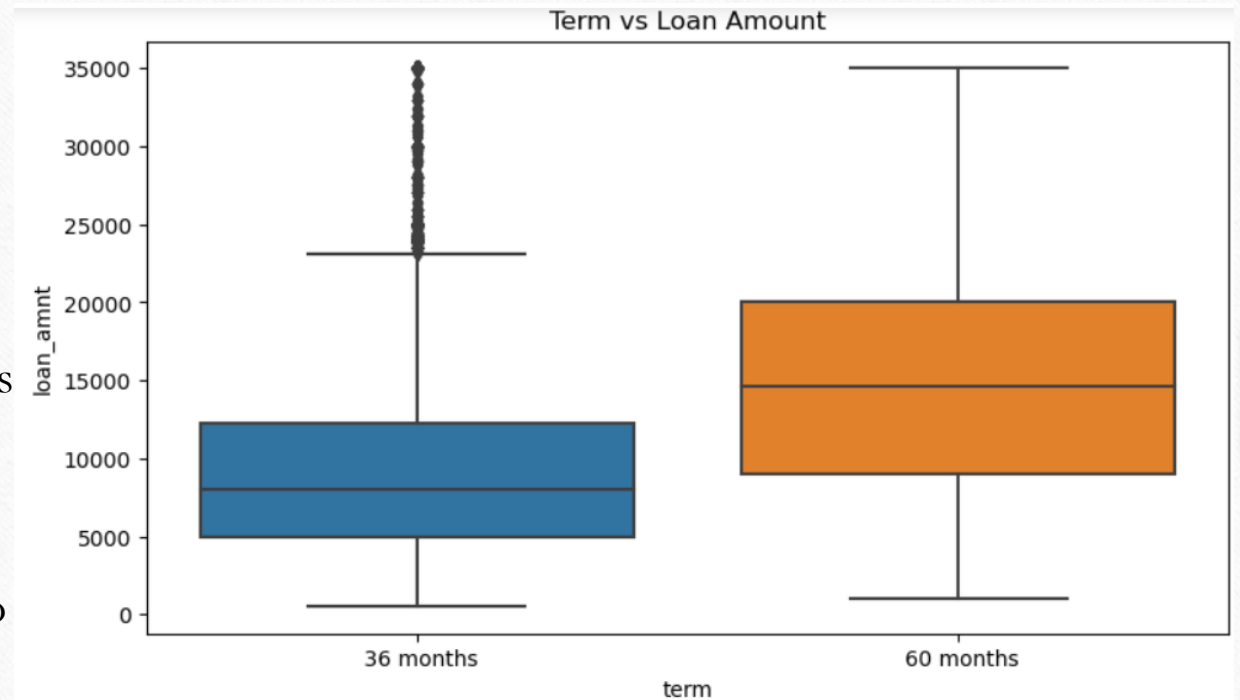
- 1) Here we can notice that highest loan amount is taken by the customers whose 'home\_ownership' is 'MORTGAGE'



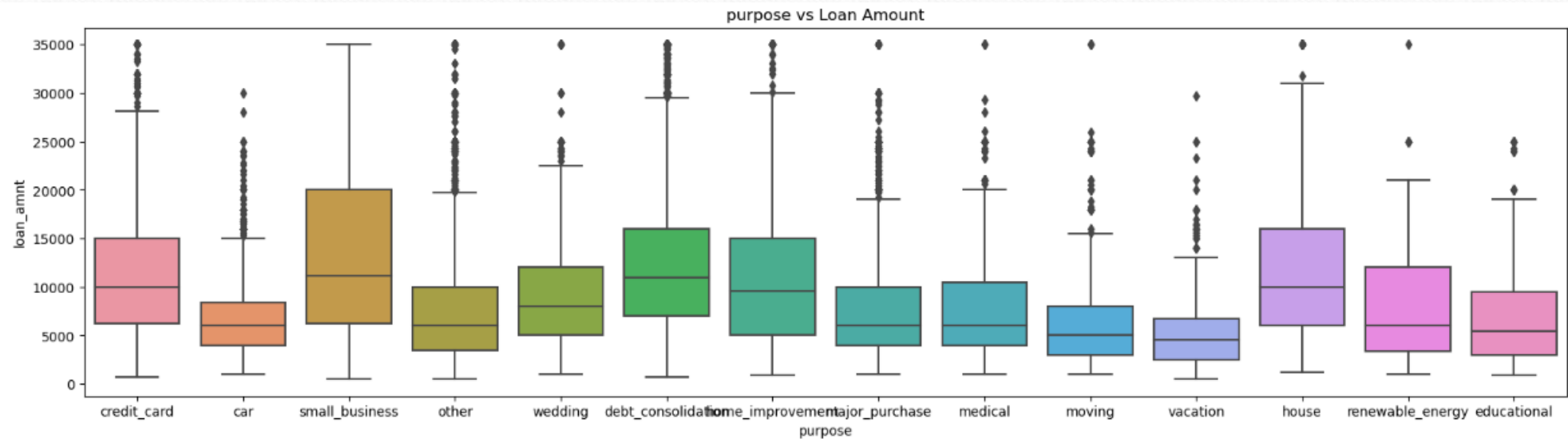
# Cont..

## Analysis on Loan Amount and Term:

1. Here we can notice that maximum loan amount dispersed is for the term of '60 months'.
2. Since the loan amount given in 36 months duration is less compared to 60 months the chance of customer getting default is low because as the tenure increase the EMI reduces and customers can afford to pay back the loan



# Cont..



Analysis on Loan Amount and Purpose:

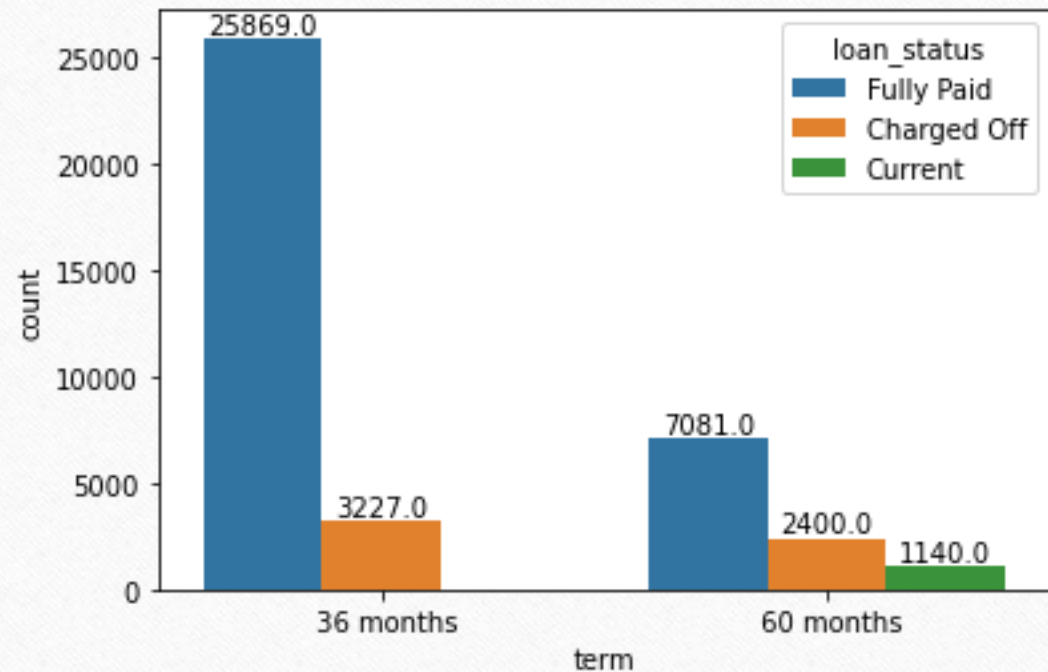
1. Here we can notice that maximum loan amount dispersed is for the purpose of 'Small Business'



# Bivariant Analysis

## 1) Analysis on Term vs Loan Status:

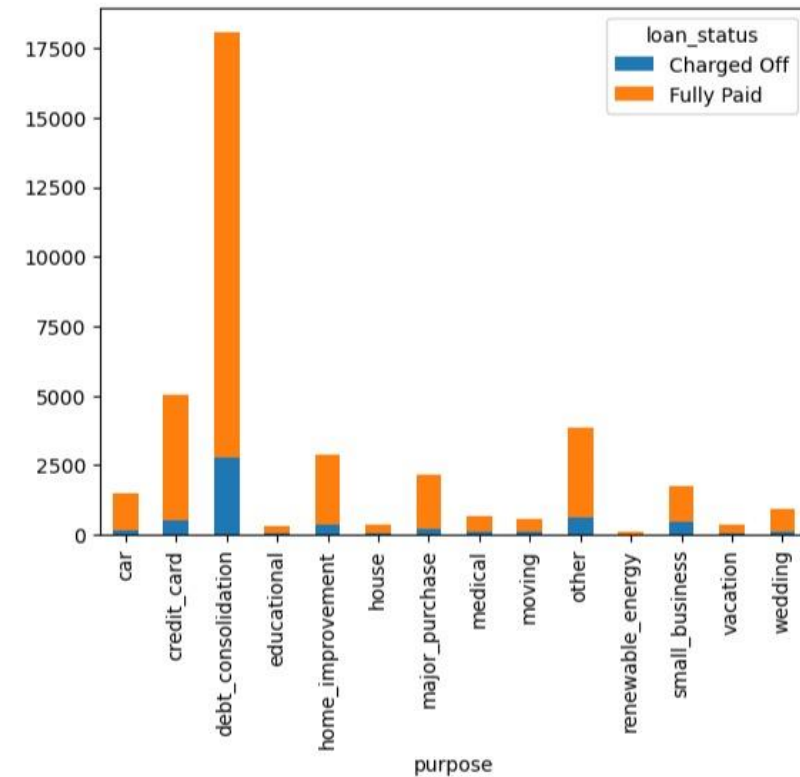
73% borrower preferred 36 months term, out of which 89% fully paid  
29% borrowers preferred 60 month tenure, out of which 60% fully paid  
Charged off is 11% of the total 36 months term loans where the same is 23% in 60 month term. Which gives an view that 60 month term loans have double burden of charging off loans



# Cont..

## 2) Analysis on Loan Status and Purpose:

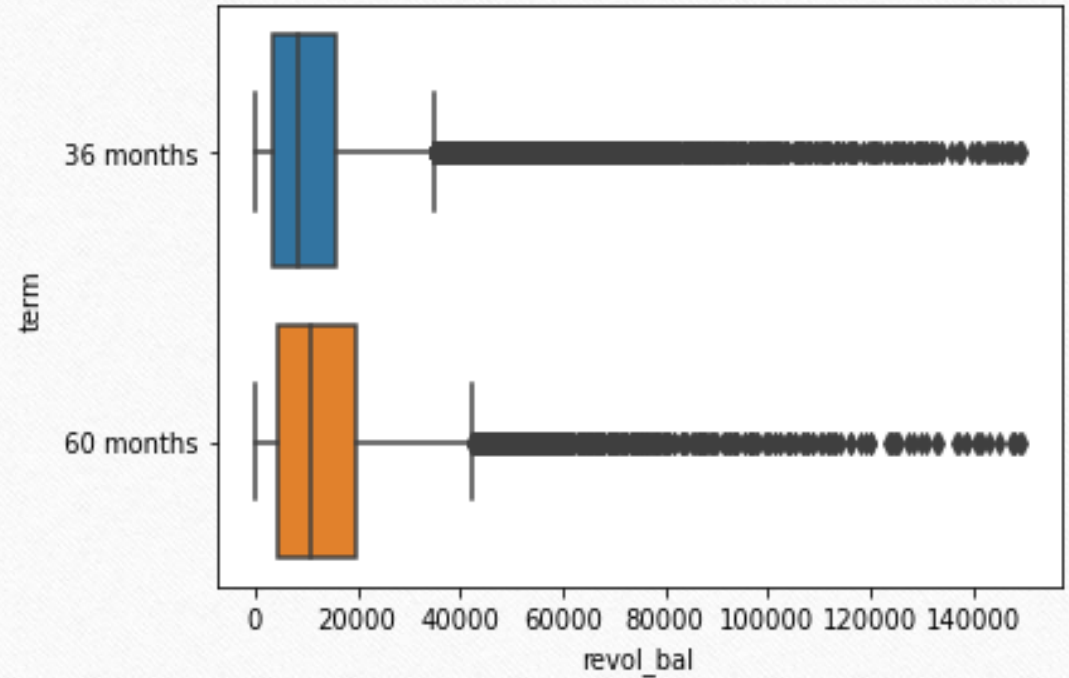
1. The stack chart shows that 49% of customers who are defaulted have purpose as 'debt consolidation'
2. This 49% debt consolidation amounts to 53% of the charged off revolving balance



# Cont..

Analysis on Term and Revolving balance:

1. The boxplot shows that for both the terms( 36 month and 60 month) maximum revolving balance confined almost in the same range, although there are few outliers present in the data set.

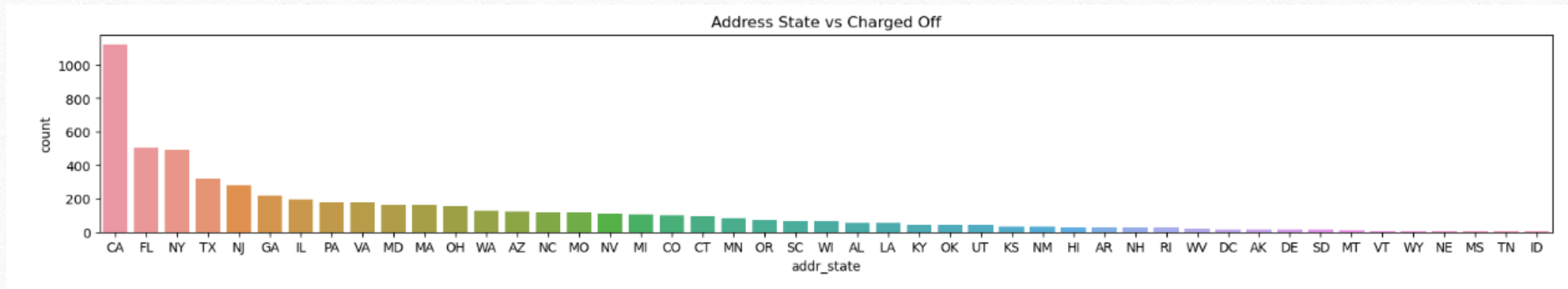




# Cont..

Analysis on Address State vs Charged Off:

1. Customers whose address state is 'CA' have more number of defaulters.
2. In the same pattern customers from states 'FL','NY', 'TX' and 'NJ' have more defaulters than other states.



# Final Insights:

---

1. Lending company should bring up a new term between 36 and 60 months which helps in reducing the defaulters
2. Verifying the applicant details before approving the loan can help in reducing the defaulters.
3. Lending company should reduce the loan dispersal to customers whose purpose is 'Small Business' to reduce the loan defaulters.
4. Lending company should look into the states where there are more defaulters exists like 'CA' , 'FL','NY', 'TX' and 'NJ'
5. Lending Company should reduce the highest loan amount dispersal to customers whose home ownership is 'MORTGAGE'.
6. Lending Company should control the loan dispersals to customers whose home ownership is 'Rent'